Scientists for Labour

Deepfakes, Elections and National Security

February 2024

Sanjush Dalmia

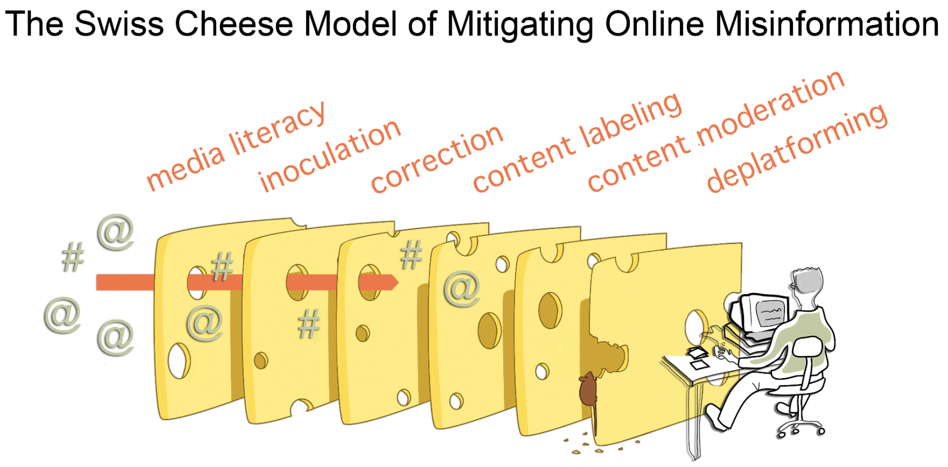
“*While policy around emerging technologies like AI has largely focussed on existential risks, the potential of AI-powered deepfakes and disinformation require urgent national and global co-operation. With high-stakes elections in over 50 countries in 2024, including the UK and the USA, we’re already behind the curve in understanding and mitigating this very real and present threat. AI-powered deep fakes can result in personalised fake news during elections – with the potential to critically undermine our trust in society and our democratic institutions. This report is a valuable contribution to the policy landscape and I hope it prompts increased attention by policymakers to develop and enact solutions fast.*”

- Daniel Aldridge MBCS, Head of Policy and Public Affairs, BCS, the Chartered Institute for IT

Introduction

Current evidence does not support the prospect of a single policy being uniquely effective in protecting elections against deepfakes.

Protection against deepfakes will likely be achieved by the additive effects of policy interventions across the development, release and use of synthetic media production tools, and the dissemination of synthetic media. This would be in line with the “Swiss Cheese Model” used to improve safety in cybersecurity, aviation and healthcare.



Leticia Bode and Emily Vraga, Bulletin of Atomic Scientists, 2021.[1]

Policy Recommendations

1. Require businesses producing synthetic media tools to demonstrate to Ofcom that media generated by their tools is identifiable at a high level of accuracy (eg - 99% or greater) by the best available deepfake detection tools, **or face fines**.

[1] The layered, Swiss cheese model for mitigating online misinformation - Bulletin of the Atomic Scientists

This market-shaping approach would a) incentivise the development of deepfake detection tools and b) incentivise the production of synthetic media which is easier to detect, via watermarks, labels and "radioactive" training data.[2] This would also align with the recommendation by EuroPol for deepfake regulations to be technology-agnostic, preventing policymakers from having to "play catch-up" as technologies advance.[3]

2. Criminalise the creation and sharing of deepfakes where *intent to deceive*[4] can be proven (as opposed to intent to entertain or educate), as a deterrent. The Online Safety Act passed in 2023 specifically criminalises the sharing of pornographic deepfakes.[5]

3. Exploit the government's convening power to accelerate the adoption of content provenance standards across industries, such as those developed by the Coalition for Content Provenance and Authority, to aid identification of real media.

4. Require creators of synthetic media to label content, in line with incoming EU legislation.[6]

5. Require large social media platforms to detect and label deepfakes, whether images, videos or audio, or **face fines** from Ofcom. Research by the Royal Society demonstrates that labels improve users' ability to identify deepfakes.[7]

6. Offer targeted support to British tech start-ups developing detection, labelling and attribution tools for deepfakes, from Innovate UK's existing budgets.

[2] Deepfakes: A Grounded Threat Assessment | Center for Security and Emerging Technology
[3] Malicious Uses and Abuses of Artificial Intelligence | Europol
[4] Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security — California Law Review
[5] Age checks, trolls and deepfakes: what's in the online safety bill?
[6] Regulating Deep Fakes in the Artificial Intelligence Act
[7] Do Content Warnings Help People Spot a Deepfake? Evidence from Two Experiments

7. Make improved "truth discernment" (the ability to correctly identify both real media and synthetic media) an explicit aim of digital literacy education, as part of Labour's "*Break down barriers to opportunity*" Mission.

8. Commit to supporting efforts at the AI Safety Institute to develop technical tools to tackle deepfakes, such as the proposed creation of a "Deepfake Zoo", a database of synthetic media to support the development of detection tools.[8]

9. Offer targeted funding to independent fact-checking services, to support truth discernment online.

10. Advocate for similar deepfake regulations at the international level, for example via David Lammy's proposed UK-EU Security Pact or via the UN, to protect British elections from foreign interference.

The Problem

In 2023 alone:

- Deepfake audio clips of Keir Starmer[9] and Sadiq Khan[10] circulated on social media.
- American voters received phone calls containing deepfake audio of Joe Biden, asking them not to vote in the New Hampshire primaries.[11]
- Two days before the Slovakian general election, a deepfake audio clip of an opposition leader appearing to be plotting to rig the election circulated on social media.[12]

The Political Deepfakes Incident Database catalogues many more examples of deepfakes being used to target politicians, potentially with the intention to influence election outcomes.

[8] Deepfakes: A Grounded Threat Assessment | Center for Security and Emerging Technology
[9] Keir Starmer deepfake shows alarming AI fears are already here - OECD.AI
[10] Sadiq Khan warns deepfakes are a 'slippery slope' after being targeted by rogue audio - OECD.AI
[11] New Hampshire investigating fake Biden robocall meant to discourage voters ahead of primary
[12] Slovakia's Election Deepfakes Show AI Is a Danger to Democracy | WIRED UK

There is also concern that an increased prevalence of deepfakes will enable politicians to raise doubts over real media which presents them in a negative light, evading accountability for their actions. This has been termed “the liar’s dividend” in political science research.[13]

Rapid advances in AI may cause deepfakes to become more persuasive, cheaper and easier to create.

The Five Missions

These policies will help Labour achieve the secure foundation of national security on which to build the Five Missions, and will help deliver the “*Take Back Our Streets*” mission.

Value for Money

With the exceptions of providing support to startups developing deepfake detection tools and to independent fact-checking services, **all of these policy interventions would be cost-free**.

The Status Quo

The Conservative government passed the Online Safety Act 2023 which criminalises the sharing of deepfake pornography,[14] but has not yet commented on the steps it is taking to protect elections against deepfakes.

A 2019 paper by DCMS concluded that we were yet to see a convincing deepfake of a politician.[15] Presumably DSIT would no longer come to this conclusion given examples of political deepfake incidents in 2023.

[13] Deepfakes, Elections, and Shrinking the Liar’s Dividend | Center for Security and Emerging Technology
[14] Age checks, trolls and deepfakes: what’s in the online safety bill?
[15] Snapshot Paper - Deepfakes and Audiovisual Disinformation - GOV.UK

In 2020, the Conservative government did not commit to any regulations in response to a DCMS committee recommendation around deepfakes.[16]

Public Opinion

In YouGov polling for ControlAI, there was 81% **net** support for “Preventing AI from being used for impersonations using the likeness or voice of people in a video, image, or sound form, without that person’s consent”.[17]

Working with Industry

TechUK has highlighted the important role the British tech industry can play in protecting elections against deepfakes, including by adopting content provenance standards and developing detection tools.[18]

Setting out Labour’s plans early with a technology-agnostic approach with offer the private sector the certainty to invest in synthetic media start-ups, and adopt synthetic media production tools for positive applications.

Learning from Abroad

The EU AI Act, which is still being deliberated upon, is likely to require creators of synthetic media to label content.[19]

In December 2023, South Korea passed legistation requiring deepfakes to be labelled and banning the sharing of political deepfakes within the 90 days before a general election.[20]

[16] Government Response to the Digital, Culture, Media and Sport Select Committee Report on Immersive and Addictive Technologies
[17] https://d3nkl3psvxxpe9.cloudfront.net/documents/ControlAI_AI_231019.pdf
[18] Deepfakes and Disinformation: What impact could this have on elections in 2024?
[19] Regulating Deep Fakes in the Artificial Intelligence Act
[20] https://www.koreatimes.co.kr/www/nation/2024/02/113_364513.html

In 2019, California banned the circulation of deepfake videos of politicians within 60 days of an election. There are concerns about the extent to which this is enforceable, but it may have a deterrent effect anyway.[21]

Differential Technology Development

“Differential technological development” refers to a governance approach which tackles risks from innovation by accelerating the development of “defensive” technologies relative to other technologies.[22]

Existing detection, labelling and attribution tools for deepfakes may become ineffective against new synthetic media production tools. Labour should aim to accelerate advances in the detection, labelling and attribution of deepfakes relative to advances in the production of synthetic media.

Playing Catch-up with Technology

Social media platforms being flooded with a high volume of persuasive deepfakes for a relatively short period of time may *irreparably* damage trust in media. Compared to the regulation of previous emerging technologies such as social media, the costs of playing catch-up are *uniquely high* with deepfakes. For this special case, Labour should err on the side of stricter regulation and relax them if needed in the future. Providing Ofxom with the flexibility to adjust accuracy requirements for the detection of synthetic media would enable regulation to be finetuned in the future.

Fraud and “the Deepfake Defence”

[21] California Looks to Boost Deepfake Protections Before Elections

[22] Differential technology development: An innovation governance consideration for navigating technology risks

In addition to election interference and deepfaked pornography, deepfakes are being used for fraud[23] and by suspected criminals to dispute CCTV evidence[24]. Deepfake regulations could therefore help Labour achieve its "*Take back our streets*" Mission.[25]

Russian Interference

The 2020 "Russia Report" found that Russia was exerting influence in British politics.[26] Britain's support for Ukraine may incentivise further interference by Russia, which could involve the use of deepfakes.

International Leadership

Building on the AI Safety Summit, advocating for deepfake regulation at the international stage will further Britain's international leadership on AI, and align with David Lammy's goal of building resilience to 21st century threats.

[23] AI: Why the next call from your family could be a deepfake scammer - BBC Science Focus Magazine

[24] The Deepfake Defense—Exploring the Limits of the Law and Ethical Norms in Protecting Legal Proceedings from Lying Lawyers

[25] Labour calls for national fraud strategy as annual losses hit £1.3bn | Scams | The Guardian

[26] The Russia report: key points and implications